INTRAH
English and French
Glossary of Training Evaluation Terms

Program for International Training in Health (INTRAH)

School of Medicine

University of North Carolina at Chapel Hill

208 North Columbia Street

Chapel Hill, North Carolina 27514 USA

1992

This publication was produced by INTRAH under United States Agency for International Development contract #AID / DPE-3031-Z-00-9024. The views expressed in this document are the responsibility of INTRAH and do not represent the policy of the U.S. Agency for International Development.

ISBN 1-881961-01-X

Table of Contents

Page

Authors

Constance Newman, M.S.W., M.P.H.
INTRAH Evaluation Officer
School of Medicine, University of North Carolina
Chapel Hill, North Carolina

Johanna Birckmayer, M.P.H.
INTRAH Graduate Assistant, July 1991-June 1992

Acknowledgements

The authors wish to express their appreciation to the following contributors:

Ms. Lynn Knauff, INTRAH Deputy Director, for providing encouragement and consultation at all stages in the development of the glossary, as well as review of all terms within the INTRAH training and program context.

Mr. Saul Helfenbein, Region Chief, Asia / Near East Division, of the Management Sciences for Health Family Planning Management Development Project, and **Dr. James Veney**, Deputy Director of the Carolina Population Center's Evaluation of Family Planning Program Impact Project, for their extensive technical review of the first draft.

Dr. Joanne Earp, Associate Professor, of the University of North Carolina School of Public Health, for her critique of several research-related definitions.

Ms. Catherine Murphy, INTRAH Training Materials Officer, for providing technical guidance throughout all production phases.

INTRAH's regional offices in Nairobi and Lomé—**Miss Pauline Muhuhu**, **Mrs. Jedida Wachira**, **Mr. Adolf Kapinga**, **Mrs. Stembile Matatu**, **Mrs. Grace Mtawali, Mr. Bongwele Onanga**, **Dr. Aliou Boly** and **Dr. Manuel Pina**—for content and language review of the second draft.

Dr. James Lea, INTRAH Director, and **Dr. Hélène Eisenberg**, INTRAH translator, for English language editing and Dr. Eisenberg for French translation. **Dr. Alain Fontaine**, INTRAH consultant, for technical review of the French text.

Ms. Jennifer Lewis, former Graduate Assistant at INTRAH, for design and desktop publishing.

Ms. Maureen Corbett, INTRAH Program Officer, for coordinating part of the French language review.

Foreword

INTRAH works with many types of health personnel in planning, conducting and evaluating family planning training. The education and backgrounds of these health workers vary tremendously, yet their responsibilities require them to share a basic knowledge of the terminologies used in family planning and primary health care training and evaluation.

In 1987, INTRAH published a bilingual *Glossary of Family Planning Terms* as a means to bring greater clarity and standardization in the language used by service providers, trainers, teachers and others involved in family planning. This bilingual *Glossary of Training Evaluation Terms* has a similar purpose; that is, to define and to promote standardization of terms frequently used by INTRAH in training evaluation at the trainee performance, activity, project and program levels.

Glossary is intended to be used as a reference by health personnel involved in training evaluation, including INTRAH staff and consultants, host country trainers and trainees, and those who sponsor or collaborate in INTRAH training and training evaluation. The definitions are concise, and the reader is assumed to have some background in training and / or evaluation. The language, aimed at a middle-level of comprehension, attempts to avoid unnecessary abstraction in an already abstract field. Those desiring a more comprehensive understanding of the terms may refer to the bibliography.

Training evaluation borrows concepts and terms from a variety of fields, including social science research, educational evaluation, statistics, performance technology and program evaluation. Consequently, the terms included in this glossary are often used in these other fields and may have specific meanings in those fields. **In this glossary, terms have been defined in the context of family planning training evaluation, with illustrative examples provided from the field of family planning training evaluation.**

The authors hope that a broad audience involved in training and training evaluation will find the *Glossary* helpful in strengthening and supporting the very important task of training evaluation.

List of Terms

English Term	French Equivalent
acceptable level of performance (ALP)	niveau acceptable de performance (NAP)
achievement	maîtrise atteinte (des objectifs)
adequacy	adéquation
appropriate	approprié
assessment	appréciation
baseline data	données de base
before and after design	plan d'enquête avant-après
biodata form	formulaire d'informations biographiques
case design	étude de cas
cause-effect relationship	relation de cause à effet
checklist	liste de référence
clinic needs assessment	recensement des besoins cliniques
closed-ended question	question fermée
cluster sample	échantillon par grappes
comprehensive training needs assessment	recensement èlargi des besoins en formation
conclusions	conclusions
confirmation	confirmation

English	French
content analysis	analyse de contenu
criterion	critère
critical tasks	tâches critiques
data	données
data analysis	analyse des données
dependent variable	variable dépendante
effectiveness	efficacité
effects	effets
efficiency	efficience
evaluation	évaluation
evaluation plan	plan d'évaluation
experimental design	plan expérimental
findings	résultats
formative evaluation	évaluation formative
goal	but
graph	graphique/diagramme
hierarchy of impacts	hiérarchie des impacts
impact	impact
independent variable	variable indépendante
indicator	indicateur

English	French
information	informations
inputs	ressources/intrants
instrument	instrument
interpretation	interprétation
intervening variable	variable intermédiaire
interview	interview
item analysis	analyse d'item
job analysis	analyse de poste
job description	description de poste
knowledge	connaissances
mean	moyenne
measure	mesure
measurement	mesure
median	médiane
methods	méthodes
mode	mode
monitoring	suivi d'exécution
multiple choice question	question à choix multiples
needs assessment	recensement/analyse des besoins
non-experimental design	plan non-expérimental

English	French
norm	norme
objective	objectif
objective test item	item de test objectif
observation	observation
open-ended question	question ouverte
operations research	recherche opérationnelle
outputs	produits/extrants
participant reaction form	questionnaire d'évaluation pour participants
performance	performance
performance analysis	analyse de la performance
performance evaluation	évaluation de la performance
population	population
post-test	post-test
pre-test	pré-test
probe (question)	question non-directive
process	processus
process review	revue du processus
progress	progrès
project review	revue du projet
purposive sample	échantillon raisonné

English	French
qualitative evaluation methods	méthodes qualitatives d'évaluation
quantitative evaluation methods	méthodes quantitatives d'évaluation
questionnaire	questionnaire
random sample	échantillon aléatoire
rating scale	échelle de notation
recommendations	recommandations
relevance	pertinence
reliability	fiabilité
research	recherche
sample	échantillon
significance (statistical)	signification (statistique)
significant difference	différence significative
skills	compétences
standard	standard
standard deviation	écart type
stratified sample	échantillon stratifié
study design	plan d'étude
summative evaluation	évaluation sommative
survey	enquête

English	French
t-test	test du t de Student
table	tableau
task analysis	analyse de tâche
time series study design	analyse en série chronologique
trainee follow-up	suivi des participants
training activity needs assessment	recensement des besoins en activités de formation
training evaluation	évaluation de la formation
validity	validité
variable	variable
verification	vérification

English Glossary
of
Training Evaluation Terms

acceptable level of performance (ALP)

a test measure that expresses the minimum acceptable performance level for competent execution of a job or task. In general discussion, the ALP is sometimes called the passing or cut-off score. It may be established by trainers, the organization or professional societies. There are many approaches used to arrive at an ALP, but in practice the ALP is often set between 65% and 75% of the maximum score. (See also **standard**.)

achievement

a) degree of proficiency of individual performance. Trainers may give pre- and post-training tests to measure trainees' achievement levels in clinical skills **b)** the degree of accomplishment of objectives of an activity, a project or a program. An evaluation activity (such as a project review) may measure the achievement of project objectives in order to evaluate project performance.

adequacy

refers to the degree to which the inputs and processes used in a program, project or activity are sufficient to meet a defined need. *Ex. The adequacy of a training program refers to whether the time and materials (inputs) spent on topics and the training activities (processes) are sufficient to meet the identified needs of trainees.* (See also **inputs**, **process**.)

appropriate

refers to the consistency of inputs, processes or outputs with the technical, cultural or situational context of a program, project or activity. *Ex. When evaluating whether a trainer is appropriate for a given workshop, a training coordinator would assess whether the trainer's skill and language profiles (inputs) were consistent with the*

content to be covered in training activities (processes) and the expected level of trainees' skill mastery (outputs). (See also **inputs**, **processes**, **outputs**.)

assessment

in general discussions, assessment is often used interchangeably with evaluation. However, in the field of educational testing, assessment refers only to the process of making observations and measurements and excludes making judgments based on the results of these measures. (See also **evaluation**.)

baseline data

data on selected indicators which describe the status of a program or project or the performance of subjects, collected prior to or at the beginning of an intervention. Baseline data may be used to set objectives, choose intervention strategies and/or may be compared with data from interim or final assessments to assess changes. *Ex. Baseline data for a training program may include: the number of FP service points, the number of trained service providers and the number of FP users.*

before and after design

a study design which involves collecting and comparing data from one or more groups before and after an intervention to determine if a change occurred. In training programs, before and after data are frequently collected using pre/post-tests, interviews and observations of performance. (See also **study design**.)

biodata form

a standardized form for collecting information on trainees' backgrounds, experience and work settings. The form is completed by trainees prior to training. The information collected may be used to

confirm that the participants' background and experience are consistent with selection criteria. The biodata form can also be used as the basis of a trainee documentation system.

case design

a non-experimental study design used to answer evaluation questions which ask for a description of the intervention, participants, goals, activities, environment, results, etc. (See also **study design**, **non-experimental design**.)

cause-effect relationship

an association between two variables in which a change in one variable, the presumed cause, produces a change in the other variable, the effect. Three criteria must be met to establish a cause-effect relationship between two variables: 1) the presumed cause should precede the effect; 2) a change in the presumed cause should be associated with a proportional change in the effect; and 3) other possible causes of the effect must be ruled out. (See also **independent variable**, **dependent variable.**)

checklist

an instrument used for collecting data. Tasks, behaviors or conditions to be evaluated are listed, and interviewers or observers determine whether the task, behavior or condition is or is not observed or present. See example on the next page. (See also **instrument**.)

checklist **Checklist for health post supervisor.**

Instructions: Read each statement below and check yes or no whether the statement describes the behavior of the person you are evaluating.

Tasks	**no**	**yes**
1. prepares monthly reports		
2. prepares the monthly plan of supervision		
3. prepares the monthly plan for the EPI program		
etc.		

clinic needs assessment

a type of needs assessment conducted to establish baseline data for a clinic site. The assessment includes the physical set-up, equipment, supplies, services, personnel, and other important features. It is used to guide project development, training, and/or practicum site identification. It is usually part of a comprehensive training needs assessment. (See also **comprehensive training needs assessment**.)

closed-ended question

a question for which there is a pre-determined or limited number of responses. *Ex. Are you satisfied with the duration of this workshop?* (See also **open-ended question**.)

cluster sample

a sample chosen by dividing the population into well-defined groups (clusters). Clusters may be villages, classes, clinics, houses, or any well-defined group of people or objects. A sample of clusters is chosen, and within each chosen cluster, additional samples can be taken. Cluster samples are often chosen for administrative convenience if a population is spread out over a large geographical area. *Ex.*

In order to obtain a sample of family planning workers, health clinics were identified as the cluster. A sample of clinics were chosen, and all workers in the clinics chosen were interviewed. (See also **sample**.)

comprehensive training needs assessment
a type of needs assessment which examines the availability, accessibility, acceptability, and quality of services; training capability and capacity; and the adequacy and content of national service policy, service standards, service plans and service protocols/procedures. A purpose is to identify training and training support needs related to the service system. (See also **needs assessment**.)

conclusions
the end result and summation of activity, project, program or study findings. *Ex. On the basis of the findings that IUDs were available but not used and the providers' statements that they did not know how to insert IUDs, it may be concluded that clients will not be able to obtain IUDs until providers are trained in IUD insertion.* (See also **findings**, **recommendations**.)

confirmation
the process of providing additional information to support the validity of the findings or conclusions of an evaluation, operations research study or needs assessment. *Ex. When evaluating a training program, the trainer should review participant reaction forms and process review results in addition to pre/ post-test results in order to confirm the validity of conclusions about the effectiveness of the training activity.*

content analysis
procedures for analyzing qualitative data, usually collected from interviews, observations or document reviews. Content analysis is conducted to organize and simplify complex data into meaningful

themes, trends or patterns. Data are first grouped according to recurring observations or statements. Then the groupings are examined for patterns, themes and trends.

criterion

a characteristic, concept or property that is examined when making a judgment about a performance, activity, program or project. Once criteria are established, a standard defining an acceptable level or range of performance or quality must be set for each criterion. *Ex. Criteria used to evaluate the quality of training include adequacy, effectiveness, efficiency and relevance.*

critical tasks

the tasks that are important and essential for effective and efficient job performance. Critical tasks include tasks on which the worker spends a high percentage of time; tasks that if not performed or performed below standards could harm clients, the organization, or the program; and tasks which must meet standards set by law or regulations. Critical tasks are identified during job analyses and may be used to set training objectives or to develop performance evaluation instruments.

data

a set of recorded observations made during needs assessments, evaluations or other research activities. Data must be analyzed and interpreted to yield usable information. (See also **information**.)

data analysis

the process of tabulating, sorting and summarizing data obtained from multiple observations in order to organize the data for interpretation. Analysis techniques can range from the simple (e.g., calculating the mean) to the more complex (e.g., t-test). (See also **data**, **interpretation**.)

dependent variable

the expected effect of a presumed cause. The dependent variable is observed in order to determine how it has responded to changes in an independent variable. *Ex. Mastery of FP skills and knowledge may be thought to depend on the duration of clinical FP training. To see if this is true, one could vary the duration of training (the independent variable) and use pre/post-test scores to measure changes in knowledge and skill levels (the dependent variable).* (See also **variable**, **independent variable**, **intervening variable**.)

effectiveness

a measure of the extent to which objectives are met. Program effectiveness refers to the extent to which the program solved the identified problem or met the identified needs of a target population or organization as expressed in the program objectives. (See also **objectives**.)

effects

the immediate or short term changes or consequences produced by a program, project or activity. *Ex. The expected effects of training 15 nurses in FP is an increase in the nurses' FP knowledge and skills and improved job performance.*

efficiency

the relationship between results (outputs) and resources (inputs). An efficiency measure estimates what it would cost in resources to achieve a unit of results. *Ex. One measure of the efficiency of a training program is the cost per participant per day.* (See also **input**, **output**.)

evaluation

the process of collecting, analyzing and interpreting data on an activity, program or project for the purpose of decision-making. The data are compared to standards or a previous status in order to make judgments about the merit or value of the program, project or activity. (See also **formative evaluation**, **summative evaluation**, **training evaluation**, **standard**.)

evaluation plan

a proposed program for the evaluation of a training project, program or activity. A project-level evaluation plan often specifies: 1) the objectives, processes and/or outcomes to be evaluated; 2) the evaluation activities, persons responsible and schedule of activities; 3) the human or financial resources necessary to carry out evaluation activities; 4) the uses to which the evaluation results will be put; 5) the means for disseminating results; and 6) any special terms of reference.

experimental design

study designs in which the investigator controls some type of intervention or other relevant factor. Experimental designs range from "classical" experimental designs, in which groups are formed by random assignment and one group receiving an intervention is compared to another group which receives no intervention or a different intervention, to quasi-experimental designs in which either randomized group selection or a control group is lacking. (See also **study design**.)

findings

a fact or facts demonstrated from the data collected during the course of an activity, project, program, or study. Findings provide the basis for conclusions and recommendations. *Ex. During a clinic needs assessment, it was observed that IUDs were not being used, although*

there was a large supply in the store room. Providers stated they did not know how to insert them. (See also **conclusions**, **recommendations**.)

formative evaluation

a type of evaluation conducted during the course of an activity, project or program. The information is used to make decisions about changes or improvements in the program, project or activity. (See also **evaluation**.)

goal

a desired, long term, general condition which a program or project can help attain. Reaching a goal is facilitated by achieving program or project objectives. *Ex. 1) Reduce population growth rate from 2.9 to 1.9 in 50 years; 2) Increase availability of quality FP services to all couples in rural areas.* (See also **objective**.)

graph

a diagram using dots, bars, lines or other symbols to visually display data. Graphs are used to display data concerning one variable or about the relationship of one variable to another. See example below showing the same data presented in two types of graphs. (See also **variable**.)

graph

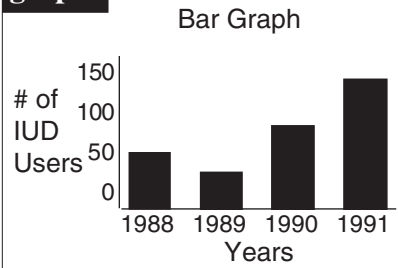

Figure 1: Number of IUD users at MCH Center "X" by year

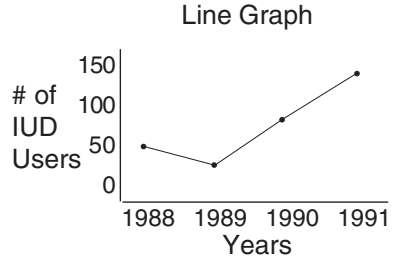

Figure 2: Number of IUD users at MCH Center "X" by year

hierarchy of impacts

a graphic presentation of the stages of a program in the form of a ladder. The hierarchy shows a logical and sequential relationship among inputs, outputs, effects and impact. It suggests that a program needs to pass through a succession of stages before it can be fully effective. (See also **inputs**, **outputs**, **effects**, **impact**.)

impact

the long term, less immediate changes or consequences produced as a result of the effects of a program or project. *Ex. An impact of training is expected to be improved quality of family planning services.* (See also **effects**.)

independent variable

the presumed cause of an effect. The independent variable is manipulated to see how it affects other dependent variables. *Ex. The duration of clinical FP training may be thought to be a cause of increased FP skills and knowledge; to test whether this is true, training duration (the independent variable) can be made shorter or longer to see if there is any variation in the mastery of FP skills and knowledge (the dependent variable.)* (See also **variable**, **dependent variable**, **intervening variable**.)

indicator

some specific, measurable aspect of an event, object or activity which can be used to give an idea about its general status or condition. Indicators are used when it is difficult or impossible to directly measure the general status. Often several indicators need to be used together to convincingly define the general status or condition. Evaluations, needs assessments and research activities must specify indicators which will be used as the basis for collecting data. *Ex. Total fertility rate (TFR) and contraceptive prevalence rate (CPR) are indicators of the effectiveness of a family planning program. The percentage of*

IUD removals within the first 3 months of insertion due to clients' reports of pelvic infections is an indicator of service provider performance.

information

the results of data analysis and interpretation, communicated in a way that can be used for decision-making. (See also **data**, **data analysis**, **interpretation**.)

inputs

the types and/or quantities of monetary, human, time and material resources used to carry out a program, project or activity. *Ex. Training inputs include financial assistance, trainers, a training facility, supplies and training materials.* (See also **outputs**, **process**.)

instrument

a form for collecting and recording data. Instruments used in training evaluation include tests, checklists, questionnaires and interview forms. (See also **data**.)

interpretation

the explanation of the meaning and significance of data analysis results. Interpretation explains the patterns, trends and relationships identified during data analysis. Objectives and standards often must be referred to in order to interpret data analysis results. (See also **data analysis**.)

intervening variable

a variable that is influenced by the independent variable and influences the dependent variable. This variable mediates between the presumed cause (independent variable) and the presumed effect (dependent variable). *Ex. In studying the relationship of training duration (the independent variable) to mastery of FP skills and knowledge (the dependent variable), an intervening variable may be trainees'*

regularity of attendance. A long training program may make it difficult for trainees to attend each day; therefore, their mastery of FP skills and knowledge will be incomplete. (See also **variable**, **dependent variable**, **independent variable**.)

interview

a method for collecting data in which one person asks questions of another person or of a group. Interviews are often used to collect data when questions are complex or sensitive, when probing questions will be asked, or when participants cannot read or write. Interviews require skilled interviewers and take more time than some other data collection methods.

item analysis

a procedure for determining the ability of a test to measure levels and changes in levels of the competence of a group of trainees in a given subject area. It is conducted to eliminate test questions which will not measure change in levels of competence. Item analysis determines the difficulty of a test item, the test item's ability to differentiate between varying levels of competence and, on multiple choice questions, the plausibility of incorrect answers, i.e., the tendency for students to select incorrect answers.

job analysis

the process of breaking down a job into smaller units including responsibilities, duties, activities, tasks and points of coordination with other work units. The results may be used to develop performance evaluation systems and to make organizational changes to improve job performance. Task analysis is one approach to job analysis. (See also **performance evaluation**, **task analysis**.)

job description

a document which provides a detailed description of the duties, responsibilities, activities and supervisory relationships for a specific job or position.

knowledge

a framework of interrelated concepts and facts that give meaning to events, support new insights and problem- solving efforts, and guide the application of skills. In training, knowledge objectives are referred to as "cognitive" objectives. *Ex. When counseling a client, a service provider must know the methods available, their advantages and disadvantages, the contraindications for each, and the client's health and social profile.* (See also **skills**.)

mean

the arithmetic average of a set of numbers. The mean is calculated by adding the numbers in the set and dividing the sum by the number of elements in the set. The mean is commonly used to summarize or describe a set of measures, such as test scores, and provides a useful indicator of group performance. The symbol for the mean is $\bar{x}$. See example below. (See also **median**, **mode**.)

mean

Trainee	Pre-test Score	
Agbli	11	Sum: 34
Tabo	9	Number of Scores: 5
Kossi	8	$\bar{x} = 34/5 = 6.8$
Didier	3	
Isaka	3	

measure
a number which provides information about an individual or group, object or activity. *Ex. A measure of a person's normal body temperature is 98.6° F. A measure of a provider's skill might be a ranking on a scale of one (inadequate) to five (adequate).*

measurement
the process of classifying or assigning values (usually numerical) to observations according to an agreed-upon procedure. The product of measurement is a quantifiable or precisely defined result.

median
the middle number of any group of numbers which have been arranged in order of magnitude. The median separates the set of numbers into 2 groups, with 50% of the measures above the median and 50% below. See example below. (See also **mean**, **mode**.)

median

Trainee	Pre-test Score	
Agbli	11	
Tabo	9	**median** = 8
Kossi	8	
Didier	3	
Isaka	3	

methods
procedures, techniques and/or processes employed during planning and implementation of evaluation activities. *Ex. There are methods to select samples, to collect data and to analyze the data collected.*

mode

the most frequently occurring number in any group of numbers. See example below. (See also **mean**, **median**.)

mode

For the numbers: 11 9 8 6 **3** **3**

mode = 3

monitoring

at the level of a program, a project or a training activity, the systematic and periodic measurement of selected indicators to determine if procedures or activities are implemented as planned and/or how far program results are deviating from expectations. The purpose of monitoring is to help trainers, supervisors or managers to use resources (inputs), execute activities on schedule, and to identify potential problems involved in obtaining desired results (outputs). (See also **inputs**, **outputs**, **progress**.)

multiple choice question

a question that requires the respondent to choose the correct or most correct response from a list of possible responses. Multiple choice questions are scored using a standard scoring key. See example below. (See also **objective test item**.)

multiple choice question

Instructions: Read the question and circle the correct response:

1. Where does fertilization normally take place?

a) in the ovary
b) in the fallopian tube
c) in the uterus
d) in the vagina

needs assessment
a systematic study of individuals, groups or organizations to identify strengths and weaknesses, and gaps between existing and optimal conditions. Data obtained from a needs assessment may be used to develop a goal and objectives, and/or they may be compared to end-of-project data in order to measure changes produced by an intervention. Needs assessments are generally conducted as one of the first steps of a program, project or activity. (See also **clinic needs assessment**, **comprehensive training needs assessment**, **training activity needs assessment**.)

non-experimental design
a study design in which an investigator collects data without intentionally intervening or controlling relevant factors. This type of study design might be used to observe, describe or document some situation or to examine some event or phenomena after the phenomena have occurred. (See also **study design**.)

norm
the expected behavior or established practice of a group. *Ex. Trainers may set what will be the practice or norm for workshop processes, such as starting on time, speaking one at a time, etc.*

objective
an expected result or accomplishment which is specific, measurable, attainable, reasonable and time-bound (SMART). The achievement of one or more objectives facilitates the attainment of a program's goal. *Ex. To provide 50% of the 44 rural area MCH centers with personnel qualified to deliver IUDs, OCs and NORPLANT® by the end of 1993.* (See also **goal**.)

objective test item
a test item for which there are standard pre-determined answers. Objective test items are scored with a standard scoring key, which does not require the scorer to make judgments about the responses; instead, the scorer must determine if the response matches the answer provided on the scoring key. True-false questions and multiple choice questions are both objective test items. (See also **multiple choice question**.)

observation
a method for collecting data which involves systematically watching and recording selected aspects of behavior and activities in the setting in which the behaviors and activities occur. Observation is often used to collect data to evaluate a subject's performance or capability.

open-ended question
a question for which there are no pre-determined answers. The respondent is able to answer in whatever way he/she chooses. *Ex. What did you think of the workshop?* (See also **close-ended question**.)

operations research
a systematic study of program activities or innovations in order to determine their effect and/or impact on program efficiency, effectiveness, or quality. Operations research is often conducted to test various approaches to solving service delivery problems. Operations research normally involves 4 steps: 1) identification and diagnosis of an operational problem; 2) formulation of a strategy to address the problem; 3) implementation and evaluation of the strategy; and 4) dissemination and application of the findings to improve operations.

outputs
the types and quantities of items produced by a program, project or activity. *Ex. An output of a FP training program for nurses is 15 trained nurses.* (See also **inputs**, **process**.)

participant reaction form
a standardized form completed by trainees at the end of training to collect data on participants' opinions about the quality of the training activity. The information is used to determine the extent to which trainees feel training objectives were met and the extent of trainee satisfaction with the training process, content, materials and trainers. Trainee satisfaction is then associated by trainers with trainee learning or failure to learn.

performance
the execution of a task. Components of performance include the knowledge and skills required to execute the task and the application of these skills and knowledge in performing the task. An individual's performance may be assessed or measured through observation. (See also **knowledge**, **skill**.)

performance analysis
the process of identifying and analyzing possible causes of job performance problems in order to develop appropriate approaches to address the problems and improve job performance. Possible approaches to performance problems include training, changes in working conditions or changes in personnel recruitment and selection. (See also **performance**.)

performance evaluation
a type of evaluation conducted at the worksite, during trainee follow-up or supervisory activities to determine if the work of service providers meets professional and organizational standards. In perfor-

mance evaluation, observed performance may be compared to pre-training performance to identify changes in the quality and regularity of performance resulting from training, or compared to a standard to determine whether performance meets expectation . If a trained service provider's performance is below standard, evaluators conduct a performance analysis to identify possible causes of performance problems, and may make recommendations as to how to improve training. (See also **performance analysis**, **standard**.)

population

the entire group which is the beneficiary of an intervention, from which a sample is drawn and to whom evaluation results are generalized. (See also **sample**.)

post-test

an instrument used to determine trainees' exit levels of knowledge or skills. A post-test is usually administered immediately after a training activity, and should be the same instrument used for the pre-test. Comparisons between pre-test and post-test results show how much participants have learned. Post-tests may be given in written form or may consist of interviews or observations of trainees' performance. (See also **pre-test**.)

pre-test

an instrument used to determine trainees' entry levels of knowledge or skill. A pre-test is administered before a training activity starts. Results may be used to identify training needs of participants and/or to provide a baseline against which post-test results or future learning may be compared. Pre-tests may be given in written form or may consist of interviews or observations of trainees' performance. (See also **baseline data**, **post-test**.)

probe (question)
verbal or non-verbal prompting used in interviews to encourage a respondent to give a more complete answer to a question or to clarify or deepen a response. *Ex. "Anything else?" "Tell me more."*

process
an operation or a succesion of operations which help or hinder the achievement of a result. A process may also consist of a planned succession of activities. *Ex. A training program (process) uses techinical or financial resources (inputs) in order to increase the number of trained service providers (outputs).* (See also **inputs**, **outputs**.)

process review
an assessment of an activity conducted in order to identify which components helped or hindered the achievement of objectives or completion of a task. *Ex. At the end of the day or session trainees were asked: 1) What went well? 2) What could be improved? 3) What could or should be done differently next time? 4) What could or should be continued next time?*

progress
refers to the extent to which the inputs, processes and outputs of an activity, project or program are occurring as planned and on schedule. Progress is monitored by tracking and comparing actual activities to planned activities. (See also **inputs**, **outputs**, **process**, **monitoring**.)

project review
a formative (annual and mid-project) or summative (end of project) evaluation whose purpose is to determine if the project is achieving the objectives and to observe and (where necessary) revise the processes employed to achieve them. The review includes assessment of both

administrative and project processes to determine if the project is being/was implemented as planned, to identify and solve implementation problems, and to plan for the future.

purposive sample
a sample chosen on the basis of a judgment about what is most useful for a study or representative of the population under study. Although purposive sampling is not valid for making generalizations to the population, the results of a study based on a purposive sample can provide important insights into the strengths and weaknesses of a program or intervention. *Ex. If the training of Community Based Distribution (CBD) workers in counseling is thought to influence method continuation rates, a study might focus on CBD projects which have innovative counseling components.* (See also **sample**.)

qualitative evaluation methods
evaluation methods which produce detailed descriptions of situations or events, or direct quotations from people about their experiences or feelings. Data about events, activities, feelings or experiences are not confined to pre-determined response categories such as "yes" or "no." Qualitative methods are particularly appropriate for the conduct of process evaluations, e.g., when the purpose of the evaluation is to develop an in-depth understanding of how a training activity, program or project produced particular outcomes or to find explanations for successes or failures in performance. An example of a qualitative evaluation method is the standardized open-ended interview.

quantitative evaluation methods
evaluation methods used to collect data on selected indicators using numbers or numerical scales. Quantitative methods are often used to examine relationships between two or more variables, to measure reactions of many subjects to a limited set of questions, or to determine how much change occurred as a result of a training activity. A written pre/post-test of knowledge using multiple choice questions is an example of a quantitative evaluation method.

questionnaire
an instrument for collecting data that consists of a list of questions. Questionnaires can be used to collect factual data or data on respondents' opinions and feelings. The questionnaires can be filled out directly by the subject from whom information is requested or by a trained interviewer. *Ex. A participant reaction form, administered at the end of training to collect participants' opinions, is a type of questionnaire.* (See also **instrument**.)

random sample
a type of sample chosen in such a way that every member of the population has an equal chance of being selected. (See also **sample.**)

rating scale
a format for collecting data which uses a scale of graduated measures for ranking levels of achievement. Tasks, behaviors or conditions to be evaluated are listed, and interviewers or observers rate the level of performance by circling the appropriate value on the scale. See example on the next page.

rating scale

Instructions: Check the response that best describes your feelings or reactions about this aspect of the training.

1. Workshop objectives were clear and were achieved.

strongly agree	agree	undecided	disagree	strongly disagree
5	4	3	2	1

recommendations

suggested actions that should take place. Recommendations are developed based on the conclusions of an activity, project, program or study. *Ex. Based on the conclusion that providers' inability to insert IUDs is preventing clients from obtaining IUDs, it is recommended that IUD insertion be included in the next clinical training of service providers.* (See also **conclusions**, **findings**.)

relevance

refers to the extent to which an activity, project or program corresponds to identified needs. *Ex. The relevance of a training activity is assessed by examining whether the training is applicable to the work of trainees, to their experience and need, and to organizational objectives.*

reliability

the extent to which a measure or instrument consistently yields the same result when two or more investigators measure the same object, person or process. (See also **measure**.)

research

systematic investigation or experimentation using scientific methods, the results of which lead to generalizable principles. *Ex. To justify expanded IUD training of nurses, a panel of medical experts mea-*

sured the performance of 25 trained nurses and 25 physicians on several indicators of proficiency in inserting IUDs. Results of the assessment showed that the nurses exceeded the physicians' performance on every measure except time. The results of this research study led to formulation of a national policy permitting nurses to insert IUDs without medical supervision.

sample

a subset or miniature version of the population. Samples are often chosen if the population is very large or difficult to access. If possible, a sample should possess all the characteristics of the population from which it is drawn so that it is representative of the larger population. If a sample is representative, findings from the sample can be generalized to the larger population. (See also **population**, **random sample**, **stratified sample**, **cluster sample**, **purposive sample**.)

significance (statistical)

statistical significance refers to the probability that an observed relationship between 2 (or more) variables, or an observed phenomenon, is not due to chance. Generally a phenomenon or relationship is said to be statistically significant when the probability that it is due to chance is less than 5% (expressed as $p < .05$). Statistical significance should not be confused with importance. (See also **significant difference**.)

significant difference

a statistical term which refers to the probability that the difference observed between the scores (such as means) of two groups on some indicator or between the scores of one group before and after an intervention is not merely due to chance. Usually a difference is said to be significant when the probability of obtaining the difference by chance is 5% or less ($p < .05$). Statistical computations, for example

t-tests, are used to determine whether a significant difference exists. A significant difference is not the same as an important difference. *Ex. If a significant difference is found between pre/post-test scores, a trainer can have a high degree of confidence that the improvement in scores is not due to chance alone but indicates a real increase in trainees' knowledge.* (See also **t-test**.)

skills

manual, mental and interpersonal abilities to carry out the procedures, operations, methods and techniques that constitute job responsibilities. Skills need to be combined with knowledge to produce acceptable levels of performance. *Ex. When counseling a new client, a service provider must have the interpersonal abilities (skill) to communicate effectively with the client, in addition to knowledge of contraceptive methods available, method contraindications and the advantages and disadvantages of each.* (See also **knowledge**.)

standard

a minimum level or range of performance or quality considered acceptable by an organization or profession. Standards are used in evaluation to make judgments about the acceptability of performance. Standards may be set on the basis of: 1) expert opinion; 2) past performance; 3) established practices (norms); or 4) some combination of 1, 2 and 3. *Ex. The cut-off score on a test represents the standard, or acceptable level of performance, on that test.* (See also **norm**, **evaluation**.)

standard deviation

a common statistical measure used to determine how much variation exists within a set of scores, i.e., whether most scores cluster closely around the mean or whether they are widely spread out. The larger the standard deviation, the greater the variation in the scores. *Ex. The*

mean score on a pre-test was 80 with a standard deviation of 2 points. This means that approximately 95% of the scores were between 76 and 84 on the test. There is little variation in the scores, and most cluster around the mean. In another case, the mean score on a pre-test was 80 with a standard deviation of 5 points. This means that 95% of the scores were between 70 and 90 on the test. There is more variation in this set of scores. (See also **mean**.)

stratified sample

a sample chosen by first dividing the population into subgroups, or strata, based on some characteristic of the population. A sample is then chosen from each stratum. A stratified sample is used to insure that all subgroups in the population are proportionately represented, especially when a particular stratum is of special interest. *Ex. To obtain a representative sample of health workers for a training needs assessment, workers may be stratified according to their years of experience and a sample chosen from each experience group.* (See also **sample**.)

study design

a plan describing the population, sample, sampling procedure, and time frame for collecting data during the course of a study. The study design is critical to acceptance of the validity of the study and is often the basis for accepting or rejecting the results of a study. An acceptable study design should convincingly demonstrate that expected results have occurred and that the results are not due to other interventions. (See also **experimental design**, **non-experimental design**, **before and after design**, **time series design**, **case design**.)

summative evaluation

a type of evaluation conducted at the end of an activity, project or program to determine the extent to which objectives have been met. The information is often used to make decisions about whether to continue the activity, project or program being evaluated. (See also **evaluation**.)

survey

a strategy for conducting an evaluation. A survey is used to collect data about a large group, generally by means of interviews or questionnaires administered to a representative sample of the group.

t-test

a statistical test used to determine if there is a significant difference between the mean scores of two groups or between the mean scores of one group on pre/post-tests. *Ex. A trainer uses a t-test to determine if there was a significant difference between the mean scores of trainees' pre/post-tests.* (See also **significant difference**.)

table

a systematic visual arrangement of data. Tables usually consist of labelled rows and columns of numbers or text. A table should be titled and have a number to identify it. See example below.

table

Table 1: Results of Pre- and Post-tests

Trainee	Pre-test Score	Post-test Score	
Tabo	77	88	
Agbli	65	75	
Kossi	70	85	etc.

task analysis

the process of identifying the activities involved in a job, breaking these down into specific tasks and determining the skills and knowledge necessary to accomplish each task. This information is often used in preparing training needs assessment or performance evaluation instruments or in the development of a training curriculum. (See also **knowledge**, **skill**.)

time series design

a study design which involves periodic data collection on one or more indicators before, during and after an intervention. The purpose is to assess whether measurements made after an intervention are a continuation of earlier patterns, or whether they mark a change. (See also **study design**.)

trainee follow-up

an activity conducted to assess formally the effects and impact of training on job performance, optimally through visits to trainees at their worksites. Follow-up is conducted to determine the extent to which the trainee has applied skills and knowledge acquired during training at his or her worksite, to identify the conditions which promote or impede the application of learning, and to determine whether the applications improved the quality and quantity of services. The findings may be used to assess needs for future training, to make recommendations for improvements in existing training activities or workplace conditions, or to describe the impact of a particular training activity on services.

training activity needs assessment

a type of needs assessment conducted by trainers before a training activity to assess the entry-level skills and knowledge of trainees in order to set training objectives and develop appropriate activities. The

assessment is based on performance standards, task analyses and expected post-training functions. (See also **needs assessment**.)

training evaluation

an appraisal made of the value (relevance, effectiveness, adequacy) of a training activity or program, a curriculum, a particular instructional procedure, or other aspects of training. The evaluation results are used to make decisions about improvement, continuation, expansion, replication or termination of training. (See also **relevance**, **effectiveness**, **adequacy**.)

validity

the extent to which a measure actually represents (measures) the status or condition which one wants to measure. *Ex. If questions on a pre/post-test reflect course objectives and what was actually taught during the course, the test instrument may be said to have content validity.* (See also **measure**.)

variable

a characteristic that can vary and take on various qualitative or quantitative values. Variables can be measured numerically *(ex. age, height)* or categorized in terms of mutually exclusive states *(ex. sex, type of client)*. (See also **independent variable**, **dependent variable**, **intervening variable**.)

verification

the process of inspecting data to ensure that they are of the expected quality, quantity and content. Verification is conducted to screen for data problems such as: errors in coding; incomplete responses; inconsistent observations of the same behavior by two raters; or low return rates. *Ex. A questionnaire was sent to trainees to assess their pre-*

training level of knowledge. When the questionnaires were returned to the trainer, she compared them with the mailing list to verify that all trainees had returned their questionnaires, and she read each form to verify that each question was answered and the data provided sufficient information for curriculum planning. (See also **data**.)

Bibliography

Babbie E: *The Practice of Social Research*, 5th ed. Belmont CA, Wadsworth Pub Co Inc, 1989.

Baird LS, Beatty RW, Schneier CE: *The Performance Appraisal Sourcebook*. Amherst MA, Human Resource Development Press, 1982.

Berk RA: *Performance Assessment*. Baltimore, Johns Hopkins University Press, 1986.

Bernardin HJ, Beatty RW: *Performance Appraisal: Assessing Human Behavior at Work*. Boston, Kent Pub Co, 1984.

Bloom BS: *Taxonomy of Educational Objectives: Book 1, Cognitive Domain*. New York, Longman Inc, 1987.

Blumenfeld SN: *Operations Research Methods: A General Approach to Primary Health Care*. Chevy Chase MD, PRICOR, 1985.

Brindis CD, Korenbrot CC, Brown PL: *Evaluation Guidebook*. San Francisco, Center for Population & Reproductive Health Policy, University of California, 1986.

Brinkerhoff R, et al: *Program Evaluation: A Practitioner's Guide for Trainers and Educators*. Boston, Kluwer-Nijhoff Publishing, 1983.

Brownlee A, Nchinda TC, Mousseau-Gershman Y: *Health Services Research Course*. Boston, Boston University Health Policy Institute, 1983.

Case R, Andrews M, Werner W: *How Can We Do It?* New York, Interaction, 1988.

de Landsheere, G: *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris, Press Universitaire de France, 1979.

Denison OL, Rosenstock IM, Getting VA: Evaluation of Program Effectiveness. *Public Health Report* 1968;83(April):323-336.

Donabedian A: Criteria, Norms and Standards of Quality: What Do They Mean? *AJPH* 1981;71(April):409-412.

Evaluation Handbooks for Family Planning Programs. Washington DC, Family Planning Evaluation Institute, 1974.

Fisher AA, et al: *Handbook for Family Planning and Operations Research Design*. New York, Population Council, 1983.

Fitzgibbon CT, Morris LL: *How to Analyze Data*. Newbury Park CA, Sage Pub Inc, 1987.

Fitzgibbon CT, Morris LL: *How to Calculate Statistics*. Beverly Hills CA, Sage Pub Inc, 1978.

Fitzgibbon CT, Morris LL: *How to Design a Program Evaluation*. Newbury Park CA, Sage Pub Inc, 1987.

Gillings J, Douglass D: *Biostats*. Chapel Hill NC, CAVCO Publications, 1985.

Glossary of Natural Family Planning Terms. Washington DC, Institute for International Studies in Natural Family Planning, Georgetown University, 1988.

Grubb CT, Loddengaard RA: *Establishing Goals and Objectives*. Chapel Hill NC, University of North Carolina, n.d.

Guilbert JJ: *Educational Handbook for Health Personnel*. Geneva, World Health Organization, 1981.

Hannum WH, Hansen C: *Instructional Systems Development in Large Organizations*. Englewood Cliffs NJ, Educational Technical Publishers, 1989.

Health Programme Evaluation. Geneva, World Health Organization, 1981.

Herman JL, Morris LL, Fitz-Gibbon CT: *Evaluator's Handbook*. Newbury Park CA, Sage Pub Inc, 1987.

Introduction to Performance Technology: Vol 1. Washington DC, National Society for Performance and Instruction, 1986.

Jonassen DH, Hannum WH, Tessmer M: *Handbook of Task Analysis Procedures*. New York, Praeger Publishers, 1989.

Kearsley G: *Costs, Benefits and Productivity in Training Systems*. Reading MA, Addison-Wesley Pub Co, 1951.

Knauff L: *Hierarchy of Impacts and Effects of Training and Training-Related Technical Assistance Interventions* (INTRAH TIMS). Chapel Hill NC, INTRAH, 1991.

Kosecoff J, Fink A: *Evaluation Basics*. Beverly Hills, Sage Pub Inc, 1982.

Krathwohl DR, Bloom BS, Masia BB: *Taxonomy of Educational Objectives: Book 2, Affective Domain*. New York, Longman Inc, 1964.

Lecomte R, Rutman L: *Introduction aux méthodes de recherche évaluative.* Université de Carleton, Ottawa, 1982.

Lipsey MW: Theory as Method: Small Theories of Treatments, in *AHCPR Conference Proceedings, Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data*. Washington DC, US Department of Health and Human Services, n.d.

Micovic, P: *Health Planning and Management Glossary*. New Delhi, World Health Organization, 1984.

Miller RA, Fisher AA: *Operations Research Approaches to Solving Problems in the Delivery of Permanent and Long-Term Family Planning Methods*. Paper presented at the Interagency Workshop on Permanent and Long-Term Contraception, sponsored by The Association for Voluntary Surgical Contraception, Kumasi, Ghana, 1989.

Morris LL, Fitz-Gibbon CT: *How to Communicate Evaluation Findings*. Newbury Park CA, Sage Pub Inc, 1987.

Morris LL, Fitz-Gibbon CT: *How to Deal With Goals and Objectives*. Newbury Park CA, Sage Pub Inc, 1978.

Morris LL, Fitz-Gibbon CT: *How to Measure Achievement*. Beverly Hills CA, Sage Pub Inc, 1978.

Nunez J: *A New Look at Family Planning Evaluation.* Watertown MA, Pathfinder International, 1991.

Operations Research; Lessons for Policy and Programs. *Population Reports* Series J 1986;31(May-June):J813-817.

Patton MQ: *Utilization Focused Evaluation*. Newbury Park CA, Sage Pub Inc, 1986.

Phillips JJ: *Handbook of Training Evaluation and Measurement Methods*. Houston, Gulf Pub Co, 1983.

Poister T, McDavid J, Magoun A: *Applied Program Evaluation in Local Government*. Lexington MA, Lexington Books, 1979.

Poister T: *Public Program Analysis: Applied Research Methods*. Baltimore, University Park Press, 1978.

Polit D, Hungler B: *Essentials of Nursing Research: Methods and Applications.* Philadelphia, JB Lipincott Co, 1989.

Rae L: *How to Measure Training Effectiveness*. New York, Nichols Pub Co, 1986.

Reynolds J: *Beyond Evaluation*. Paper presented at the Joint Meeting of the Evaluation Research Society and the Evaluation Network, San Francisco, 1984.

Robinson D, Robinson J: *Training for Impact*. San Francisco, Jossey-Bass Publishers, 1989.

Ross JA, Donayre J, McNamara R: Perspectives on Operations Research. *International Family Planning Perspectives* 1987;13(Dec):128-136.

Rossi PH, Freeman HE: *Evaluation: A Systematic Approach*. Beverly Hills CA, Sage Pub Inc, 1985.

Russell S, Reynolds J: *Operations Research Issues: Community Financing*. PRICOR Monograph Series: Issues Paper 1. Chevy Chase MD, PRICOR, 1985.

Schaefer M: *Evaluation/Decision Making in Health Planning and Administration*. HADM Monograph Series #3. Chapel Hill NC, University of North Carolina, n.d.

Scholl EA: *Family Planning Training Evaluation Manual*. Arlington VA, Development Associates, 1988.

Scriven M: *Evaluation Thesaurus*, 4th ed. Newbury Park CA, Sage Pub Inc, 1991.

Suchman EA: *Evaluative Research*. New York, Russell Sage Foundation, 1967.

Udinsky BF, Osterlind SJ, Lynch SW: *Evaluation Resource Handbook*. San Diego, Edits Publishers, 1981.
UNESCO: *Glossary of Educational Technology Terms*. Lausanne, Presses Centrales, 1984.

Veney JE, Kaluzny AD: *Evaluation and Decision Making for Health Services Programs*. Ann Arbor, Health Administration Press, 1991.

Weiss CH: *Evaluation Research*. Englewood Cliffs NJ, Prentice-Hall Inc, 1972.

Wiersma W, Jurs SG: *Educational Measurement and Testing*. Boston, Allyn and Bacon Inc, 1985.

Windsor RA, et al: *Evaluation of Health Promotion and Education Programs*. Palo Alto CA, Mayfield Pub Co, 1984.

Wolf RM: *Evaluation in Education*. New York, Praeger Pub, 1984.

Wolff JA, Suttenfield LJ, Binzen SC(eds): *The Family Planning Manager's Handbook*. Hartford, Kumarian Press, 1991.